

CLAIMS

1. (Original) A computer-implemented method for identifying sequences of Web pages that are frequently visited in order non-consecutively during user browsing sessions, the Web pages of a sequence visited in order during a user browsing session when each Web page of the sequence is visited before a next Web page of the sequence is visited, the Web pages of a sequence visited non-consecutively during a user browsing session when at least one intervening Web page that is not part of the sequence is visited between the visits to the Web pages of the sequence, the method comprising:

receiving a Web server log generated by a Web server serving a Website having Web pages, the Web server log reflecting usage of the Website by users;

identifying multiple user browsing sessions from the received Web server log, each identified user browsing session indicating a series of Web pages from the Website that were consecutively visited in order by a user; and

identifying sequences of the Web pages of the Website that were frequently visited in order non-consecutively in the identified user browsing sessions by

determining a minimum threshold number of user browsing sessions;

determining multiple Web pages that are each visited during more of the identified user browsing sessions than the minimum threshold;

creating a tree data structure having a root node and having a child node of the root node for each of the determined Web pages, each of the determined Web pages represented by one of the children nodes, the children nodes forming a current lowest level of the tree data structure;

repeatedly expanding the tree data structure by adding a new lowest level of nodes that are children nodes to the nodes of a previous lowest level of the tree data structure, the added children nodes such that each of the determined Web pages has a node that represents that determined Web page that is added as a child node to each of the nodes of the previous level, each of the added children nodes having an associated

sequence of Web pages consisting of the determined Web pages that are represented by the nodes in a path from the root node to that node;

    determining the nodes of the tree data structure whose associated sequence of Web pages is visited in order non-consecutively during more of the identified user browsing sessions than the minimum threshold; and

    removing the nodes of the tree data structure that are not among the determined nodes,

such that after creation of the tree data structure is completed, the sequences of Web pages that are associated with the nodes remaining in the tree data structure are the identified sequences of the Web pages that were frequently visited in order non-consecutively during the identified user browsing sessions.

2. (Original) The method of claim 1 including producing a report that includes the identified frequently visited sequences of Web pages.

3. (Original) The method of claim 1 wherein the acts of determining the nodes and removing the nodes are each performed after each act of adding a new lowest level of nodes, and wherein the act of determining the nodes includes:

    initializing counters associated with each of the nodes other than the root node;

    for each of the identified user browsing sessions, incrementing the counter associated with each node whose associated sequence of Web pages is visited in order non-consecutively during the identified user browsing session; and

    after the incrementing of the counters for each of the identified user browsing sessions, selecting the nodes of the tree data structure whose associated counters have values greater than the minimum threshold to be the determined nodes.

4. (Original) The method of claim 3 including creating for each of the determined Web pages a linked list data structure that is associated with that determined

Web page, and wherein the incrementing of the counters associated with the nodes for a user browsing session includes:

for each of the linked list data structures, initializing the linked list to include an entry for the child node of the root node that represents the determined Web page with which the linked list is associated; and

selecting in order each Web page in the series indicated by the user browsing session and, if that Web page is one of the determined Web pages,

incrementing the counter associated with each of the nodes having entries in the linked list data structure associated with the one determined Web page;

selecting all children nodes of each node having an entry in the linked list data structure associated with the one determined Web page;

removing each of the entries in the linked list data structure associated with the one determined Web page; and

for each of the selected children nodes, adding an entry for the selected child node to the linked list data structure associated with the determined Web page represented by the selected child node.

5. (Original) The method of claim 3 wherein the incrementing of the counters for an identified user browsing session includes:

associating a token with the root node; and

selecting in order each Web page in the series indicated by the user browsing session and, if that Web page is one of the determined Web pages,

for each node with an associated token, determining any children nodes of that node that represent the one determined Web page; and

for each determined child node, if the node does not have an associated token,

incrementing the counter associated with the determined child node;

and

associating a token with the determined child node.

6. (Original) The method of claim 3 including, after the adding of each new lowest level of nodes but before the determining of the nodes, removing the added nodes of the new lowest level whose associated sequence of Web pages includes a subsequence of Web pages that is not associated with any of the nodes that were part of the tree data structure before the adding.

7. (Original) The method of claim 3 wherein the repeated expansion of the tree data structure is ended when all of the nodes of an added lower level are removed.

8. (Original) The method of claim 1 wherein the repeated expansion of the tree data structure is ended when a maximum number of levels of nodes is reached.

9. (Original) The method of claim 1 including, after the determining of the multiple Web pages, determining sequences of two of the determined Web pages that are visited in order non-consecutively during more of the identified user browsing sessions than the minimum threshold, wherein during the adding of each new lowest level of nodes an added child node is removed if any consecutive two of the Web pages in the sequence of Web pages associated with that child node are not one of the determined sequences of two determined Web pages, and wherein the repeated expanding of the tree data structure is completed before the determining of the nodes.

10. (Original) The method of claim 9 wherein the determining of the sequences of two of the determined Web pages includes creating an array data structure having entries representing each possible sequence of two of the determined Web pages.

11. (Original) The method of claim 1 wherein the determining of the minimum threshold number of user browsing sessions is based on an indication received from a user.

12. (Original) A computer-implemented method for identifying frequent sequences of interaction events, the method comprising:

from at least one interaction log that contains data reflecting interactions with at least one executing application, identifying multiple interaction sessions each indicating an ordered series of related interaction events; and

identifying sequences of interaction events that frequently occur in the identified interaction sessions by

determining multiple interaction events that are each present in at least a first threshold number of the identified interaction sessions;

for each of the determined interaction events, identifying a sequence consisting of that interaction event as being a sequence of interaction events of length 1 that frequently occurs, the length 1 being a current longest length of identified sequences; and

repeatedly identifying sequences of interaction events of increasing lengths that frequently occur by,

for each of the identified sequences of a current longest length, generating candidate sequences of interaction events by adding one of the determined interaction events at the end of the identified sequence;

identifying the newly generated candidate sequences that are present in at least a second threshold number of the identified interaction sessions as being sequences of interaction events that frequently occur, each of the identified newly generated sequences having a current longest length that is one greater than the previous longest length; and

if none of the newly generated candidate sequences are identified as sequences of interaction events that frequently occur, ending the repeated identification of the sequences.

13. (Original) The method of claim 12 wherein the identifying of the sequences of interaction events of length 1 includes creating a data structure with a root element and

multiple children elements of the root element, each child element representing one of the determined interaction events and being associated with the identified sequence of length 1 that consists of that one determined interaction event, the children elements forming a current lowest level of the data structure, and wherein the generating of candidate sequences of a length one greater than the current longest length includes expanding the data structure by adding a new lowest level of elements to the data structure such that the added elements are children elements to the elements of a previous lowest level of the data structure, each of the added children elements representing one of the determined interaction events and having an associated sequence of interaction events that is one of the generated candidate sequences.

14. (Original) The method of claim 13 wherein each element other than the root element has an associated sequential path of elements between the root element and that element, a first element in each sequential path being a child element of the root element, each element in each sequential path other than the first element being a child element of the previous element in the sequential path, and wherein the sequence of interaction events that is associated with each node other than the root node consists of a sequence of the interaction events represented by the elements in the path associated with that node followed by the interaction event represented by that node.

15. (Original) The method of claim 13 wherein the identifying of the newly generated candidate sequences as sequences of interaction events that frequently occur includes determining the newly added children elements whose associated sequence of interaction events is present in at least the second threshold number of the identified interaction sessions.

16. (Original) The method of claim 15 wherein the determining of the newly added children elements whose associated sequence of interaction events is present in at

least the second threshold number of the identified interaction sessions includes, for each of the identified interaction sessions:

initializing for each of the determined interaction events a linked list data structure that is associated with that determined interaction event such that the linked list data structure includes an entry for the child element of the root element that represents the determined interaction event with which the linked list is associated; and

selecting in order each interaction event in the identified interaction session and, if that interaction event is one of the determined interaction events,

indicating for each of the elements having entries in the linked list data structure associated with the one determined interaction event that the sequence of interaction events associated with the element is present in the identified interaction session;

selecting all children elements of each element having an entry in the linked list data structure associated with the one determined interaction event;

removing each of the entries in the linked list data structure associated with the one determined interaction event; and

for each of the selected children elements, adding an entry for the selected child element to the linked list data structure associated with the determined interaction event represented by the selected child element.

17. (Original) The method of claim 15 wherein the determining of the newly added children elements whose associated sequence of interaction events is present in at least the second threshold number of the identified interaction sessions includes, for each of the identified interaction sessions:

associating a token with the root element; and

selecting in order each interaction event in the identified interaction session and, if that interaction event is one of the determined interaction events,

for each element with an associated token, determining any children elements of that element that represent the one determined interaction event; and

for each determined child element, if the element does not have an associated token,

indicating that the sequence of interaction events associated with the determined child element is present in the identified interaction session; and

associating a token with the determined child element.

18. (Original) The method of claim 15 wherein the identifying of the newly generated candidate sequences as sequences of interaction events that frequently occur includes, after the determining of the newly added children elements whose associated sequence of interaction events is present in at least a second threshold number of the identified interaction sessions, removing the newly added children elements that are not among the determined elements.

19. (Original) The method of claim 12 wherein the second threshold number is the first threshold number.

20. (Original) The method of claim 12 wherein the generating of the candidate sequences is such that a candidate sequence of interaction events is not generated if the candidate sequence includes a subsequence of interaction events that is not one of the identified sequences of interaction events that frequently occurs.

21. (Original) The method of claim 12 including:

determining a maximum sequence length for the identified sequences of interaction events that frequently occur; and

ending the repeated identification of the sequences after the identifying of the newly generated candidate sequences that have a current longest length that is the specified maximum sequence length.

22. (Original) The method of claim 12 including providing indications of the identified sequences of interaction events that frequently occur to at least one user.

23. (Original) A computer-readable medium containing instructions that when executed cause a computing device to identify frequent sequences of interaction events by:

identifying multiple interaction sessions each indicating an ordered series of related interaction events; and

identifying sequences of interaction events that frequently occur in the identified interaction sessions by

determining multiple interaction events that are each present in at least a first threshold number of the identified interaction sessions;

for each of the determined interaction events, identifying a sequence consisting of that interaction event as being a sequence of interaction events of length 1 that frequently occurs, the length 1 being a current longest length of identified sequences; and

repeatedly identifying sequences of interaction events of increasing lengths that frequently occur by,

for each of the identified sequences of a current longest length, generating candidate sequences by adding one of the determined interaction events at the end of the identified sequence;

identifying the newly generated candidate sequences that are present in at least a second threshold number of the identified interaction sessions as being sequences of interaction events that frequently occur, each of the identified newly generated sequences having a current longest length that is one greater than the previous longest length; and

if none of the newly generated candidate sequences are identified as sequences of interaction events that frequently occur, ending the repeated identification of the sequences.

24. (Original) The computer-readable medium of claim 23 wherein the identifying of the sequences of interaction events of length 1 includes creating a data structure with a root element and multiple children elements of the root element, each child element representing one of the determined interaction events and being associated with the identified sequence of length 1 that consists of that one determined interaction event, the children elements forming a current lowest level of the data structure, and wherein the generating of candidate sequences of a length one greater than the current longest length includes expanding the data structure by adding a new lowest level of elements to the data structure such that the added elements are children elements to the elements of a previous lowest level of the data structure, each of the added children elements representing one of the determined interaction events and having an associated sequence of interaction events that is one of the generated candidate sequences.

25. (Original) The computer-readable medium of claim 24 wherein the identifying of the newly generated candidate sequences as sequences of interaction events that frequently occur includes determining the newly added children elements whose associated sequence of interaction events is present in at least the second threshold number of the identified interaction sessions by, for each of the identified interaction sessions:

initializing for each of the determined interaction events a linked list data structure that is associated with that determined interaction event such that the linked list data structure includes an entry for the child element of the root element that represents the determined interaction event with which the linked list is associated; and

selecting in order each interaction event in the identified interaction session and, if that interaction event is one of the determined interaction events,

indicating for each of the elements having entries in the linked list data structure associated with the one determined interaction event that the sequence of interaction events associated with the element is present in the identified interaction session;

selecting all children elements of each element having an entry in the linked list data structure associated with the one determined interaction event;

removing each of the entries in the linked list data structure associated with the one determined interaction event; and

for each of the selected children elements, adding an entry for the selected child element to the linked list data structure associated with the determined interaction event represented by the selected child element.

26. (Original) The computer-readable medium of claim 24 wherein the identifying of the newly generated candidate sequences as sequences of interaction events that frequently occur includes determining the newly added children elements whose associated sequence of interaction events is present in at least the second threshold number of the identified interaction sessions by, for each of the identified interaction sessions:

associating a token with the root element; and

selecting in order each interaction event in the identified interaction session and, if that interaction event is one of the determined interaction events,

for each element with an associated token, determining any children elements of that element that represent the one determined interaction event; and

for each determined child element, if the element does not have an associated token,

indicating that the sequence of interaction events associated with the determined child element is present in the identified interaction session; and

associating a token with the determined child element.

27. (Original) The computer-readable medium of claim 23 wherein the instructions further cause the computing device to indicate the identified sequences of interaction events that frequently occur to at least one user.

28. (Original) A computing device for identifying frequent sequences of interaction events, comprising:

an interaction event supplier component capable of identifying multiple interaction sessions each indicating an ordered series of related interaction events from at least one interaction log; and

a frequent sequence analyzer component capable of identifying sequences of interaction events that frequently occur in the identified interaction sessions by

determining multiple interaction events that are each present in at least a first threshold number of the identified interaction sessions;

for each of the determined interaction events, identifying a sequence consisting of that interaction event as being a sequence of interaction events of length 1 that frequently occurs, the length 1 being a current longest length of identified sequences; and

repeatedly identifying sequences of interaction events of increasing lengths that frequently occur by,

for each of the identified sequences of a current longest length, generating candidate sequences by adding one of the determined interaction events at the end of the identified sequence;

identifying the newly generated candidate sequences that are present in at least a second threshold number of the identified interaction sessions as being sequences of interaction events that frequently occur, each of the identified newly generated sequences having a current longest length that is one greater than the previous longest length; and

if none of the newly generated candidate sequences are identified as sequences of interaction events that frequently occur, ending the repeated identification of the sequences.

29. (Original) The computing device of claim 28 including a frequent sequence indication generator component capable of providing to at least one user indications of the identified sequences of interaction events that frequently occur.

30. (Original) The computing device of claim 28 wherein the identifying of the sequences of interaction events of length 1 includes creating a data structure with a root element and multiple children elements of the root element, each child element representing one of the determined interaction events and being associated with the identified sequence of length 1 that consists of that one determined interaction event, the children elements forming a current lowest level of the data structure, and wherein the generating of candidate sequences of a length one greater than the current longest length includes expanding the data structure by adding a new lowest level of elements to the data structure such that the added elements are children elements to the elements of a previous lowest level of the data structure, each of the added children elements representing one of the determined interaction events and having an associated sequence of interaction events that is one of the generated candidate sequences.

31. (Original) The computing device of claim 30 wherein the identifying of the newly generated candidate sequences as sequences of interaction events that frequently occur includes determining the newly added children elements whose associated sequence of interaction events is present in at least the second threshold number of the identified interaction sessions by, for each of the identified interaction sessions:

initializing for each of the determined interaction events a linked list data structure that is associated with that determined interaction event such that the linked list data structure includes an entry for the child element of the root element that represents the determined interaction event with which the linked list is associated; and

selecting in order each interaction event in the identified interaction session and, if that interaction event is one of the determined interaction events,

indicating for each of the elements having entries in the linked list data structure associated with the one determined interaction event that the sequence of interaction events associated with the element is present in the identified interaction session;

selecting all children elements of each element having an entry in the linked list data structure associated with the one determined interaction event;

removing each of the entries in the linked list data structure associated with the one determined interaction event; and

for each of the selected children elements, adding an entry for the selected child element to the linked list data structure associated with the determined interaction event represented by the selected child element.

32. (Original) The computing device of claim 30 wherein the identifying of the newly generated candidate sequences as sequences of interaction events that frequently occur includes determining the newly added children elements whose associated sequence of interaction events is present in at least the second threshold number of the identified interaction sessions by, for each of the identified interaction sessions:

associating a token with the root element; and

selecting in order each interaction event in the identified interaction session and, if that interaction event is one of the determined interaction events,

for each element with an associated token, determining any children elements of that element that represent the one determined interaction event; and

for each determined child element, if the element does not have an associated token,

indicating that the sequence of interaction events associated with the determined child element is present in the identified interaction session; and

associating a token with the determined child element.

33. (Original) A computer-implemented method for identifying frequent sequences of interaction events, the method comprising:

from at least one interaction log that contains data reflecting interactions with at least one executing application, identifying multiple interaction sessions each indicating an ordered series of related interaction events; and

identifying sequences of interaction events that frequently occur in the identified interaction sessions by

determining sequences of interaction events each having a specified number of multiple interaction events such that the determined sequences are each present in at least a first threshold number of the identified interaction sessions;

generating candidate sequences of interaction events having varying lengths, each generated candidate sequence such that each subsequence of the candidate sequence that is of a length that is the specified number is one of the determined sequences, by adding at least one interaction event at the end of each of the determined sequences and by repeatedly generating additional sequences by adding at least one interaction event at the end of previously generated sequences; and

identifying the generated candidate sequences that are present in at least a second threshold number of the identified interaction sessions as being sequences of interaction events that frequently occur.

34. (Original) The method of claim 33 wherein the generating of the candidate sequences of interaction events includes creating a multi-level data structure having a root element and a plurality of other elements, a first of the levels of the data structure having elements that are children elements of the root element, each of the other levels having elements that are children elements of elements of a previous level, and each of the other elements representing one of the interaction events and being associated with one of the generated candidate sequences.

35. (Original) The method of claim 34 wherein each element other than the root element has an associated sequential path of elements between the root element and that element, a first element in each sequential path being a child element of the root element, each element in each sequential path other than the first element being a child element of the previous element in the sequential path, and wherein the sequence of interaction events that is associated with each node other than the root node consists of a sequence of the interaction events represented by the elements in the path associated with that node followed by the interaction event represented by that node.

36. (Original) The method of claim 34 wherein the identifying of the generated candidate sequences as being sequences of interaction events that frequently occur includes determining the elements of the data structure whose associated sequence of interaction events is present in at least the second threshold number of the identified interaction sessions.

37. (Original) The method of claim 36 wherein the determining of the elements of the data structure whose associated sequence of interaction events is present in at least the second threshold number of the identified interaction sessions includes, for each of the identified interaction sessions:

for each interaction event represented by a child element of the root element, initializing a linked list data structure that is associated with that interaction event such that the linked list data structure includes an entry for the child element of the root element that represents the interaction event with which the linked list is associated; and

selecting in order each interaction event in the identified interaction session and, if that interaction event is one of the interaction events having an associated linked list data structure,

indicating for each of the elements having entries in the associated linked list data structure that the sequence of interaction events associated with that element is present in the identified interaction session;

selecting all children elements of each element having an entry in the associated linked list data structure;

removing each of the entries in the associated linked list data structure; and

for each of the selected children elements that represent an interaction event that has an associated linked list data structure, adding an entry for the selected child element to that linked list data structure.

38. (Original) The method of claim 36 wherein the determining of the elements of the data structure whose associated sequence of interaction events is present in at least the second threshold number of the identified interaction sessions includes, for each of the identified interaction sessions:

associating a token with the root element; and

selecting in order each interaction event in the identified interaction session and for each selected interaction event,

for each element with an associated token, determining any children elements of that element that represent the selected interaction event; and

for each determined child element, if the element does not have an associated token,

indicating that the sequence of interaction events associated with the determined child element is present in the identified interaction session; and

associating a token with the determined child element.

39. (Original) The method of claim 36 wherein the identifying of the generated candidate sequences as being sequences of interaction events that frequently occur includes, after the determining of the elements of the data structure whose associated sequence of interaction events is present in at least the second threshold number of the identified interaction sessions, removing the elements of the data structure that are not among the determined elements.

40. (Original) The method of claim 33 wherein the second threshold number is the first threshold number.

41. (Original) The method of claim 33 including:

determining a maximum sequence length for the identified sequences of interaction events that frequently occur; and

ending the repeated generation of the additional sequences after the generating of candidate sequences whose length is the specified maximum sequence length.

42. (Original) The method of claim 33 including providing indications of the identified sequences of interaction events that frequently occur to at least one user.

43. (Original) A computer-readable medium containing instructions that when executed cause a computing device to identify frequent sequences of interaction events by:

from at least one interaction log that contains data reflecting interactions with at least one executing application, identifying multiple interaction sessions each indicating an ordered series of related interaction events; and

identifying sequences of interaction events that frequently occur in the identified interaction sessions by

determining sequences of interaction events each having a specified number of multiple interaction events such that the determined sequences are each present in at least a first threshold number of the identified interaction sessions;

generating candidate sequences of interaction events of varying lengths, each generated candidate sequence such that each subsequence of the candidate sequence that is of a length that is the specified number is one of the determined sequences, by adding at least one interaction event at the end of each of the determined sequences and by repeatedly generating additional sequences by adding at least one interaction event at the end of previously generated sequences; and

identifying the generated candidate sequences that are present in at least a second threshold number of the identified interaction sessions as being sequences of interaction events that frequently occur.

44. (Original) The computer-readable medium of claim 43 wherein the generating of the candidate sequences of interaction events includes creating a multi-level data structure having a root element and a plurality of other elements, a first of the levels of the data structure having elements that are children elements of the root element, each of the other levels having elements that are children elements of elements of a previous level, and each of the other elements representing one of the interaction events and being associated with one of the generated candidate sequences.

45. (Original) The computer-readable medium of claim 43 wherein the instructions further cause the computing device to indicate the identified sequences of interaction events that frequently occur to at least one user.

46. (Original) A computing device for identifying frequent sequences of interaction events, comprising:

an interaction event supplier component capable of identifying multiple interaction sessions each indicating an ordered series of related interaction events from at least one interaction log; and

a frequent sequence analyzer component capable of identifying sequences of interaction events that frequently occur in the identified interaction sessions by

determining sequences of interaction events each having a specified number of multiple interaction events such that the determined sequences are each present in at least a first threshold number of the identified interaction sessions;

generating candidate sequences of interaction events of varying lengths, each generated candidate sequence such that each subsequence of the candidate sequence that is of a length that is the specified number is one of the determined

sequences, by adding at least one interaction event at the end of each of the determined sequences and by repeatedly generating additional sequences by adding at least one interaction event at the end of previously generated sequences; and

identifying the generated candidate sequences that are present in at least a second threshold number of the identified interaction sessions as being sequences of interaction events that frequently occur.

47. (Original) The computing device of claim 46 including a frequent sequence indication generator component capable of providing to at least one user indications of the identified sequences of interaction events that frequently occur.

48. (Original) The computing device of claim 46 wherein the generating of the candidate sequences of interaction events includes creating a multi-level data structure having a root element and a plurality of other elements, a first of the levels of the data structure having elements that are children elements of the root element, each of the other levels having elements that are children elements of elements of a previous level, and each of the other elements representing one of the interaction events and being associated with one of the generated candidate sequences.

49-94. (Cancelled)

95. (Original) A data structure stored in a memory for use in identifying frequent sequences of interaction events, the data structure containing a root element and a plurality of other elements, a first level of the data structure having elements that are children elements of the root element, each of multiple other levels having elements that are children elements of elements of a previous level, each of the plurality of other elements representing one of the interaction events and being associated with a candidate sequence,

so that after the elements are scored while validating the data structure against groups of related interaction events, the candidate sequences that are associated with elements whose score is above a threshold can be selected as the identified frequent sequences of interaction events.

96. (Original) The data structure of claim 95 wherein each element other than the root element has an associated sequential path of elements between the root element and that element, a first element in each sequential path being a child element of the root element, each element in each sequential path other than the first element being a child element of the previous element in the sequential path, and wherein the candidate sequence of interaction events that is associated with each node other than the root node consists of a sequence of the interaction events represented by the elements in the path associated with that node followed by the interaction event represented by that node.

97. (Original) The data structure of claim 95 including, for each interaction event represented by a child element of the root element, a linked list data structure that is associated with that interaction event such that the linked list data structure includes entries for at least one of the other elements, so that as the data structure is validated against one of the groups of interaction events by selecting in order each interaction event in the group, if the selected interaction event is one of the interaction events having an associated linked list data structure, the score of each of the elements having entries in the associated linked list data structure can be incremented.

98. (Original) The data structure of claim 95 wherein at least some of the elements have associated tokens, so that as the data structure is validated against one of the groups of interaction events by selecting in order each interaction event in the group, the score can be incremented for

each element that represents the selected interaction event, that lacks an associated token, and that is a child element of an element with an associated token.